ASYMPTOTIC RESULTS ON THE LENGTH OF COALESCENT TREES

JEAN-FRANÇOIS DELMAS, JEAN-STÉPHANE DHERSIN, AND ARNO SIRI-JEGOUSSE

Abstract. We give the asymptotic distribution of the length of partial coalescent trees for Beta and related coalescents. This allows us to give the asymptotic distribution of the number of (neutral) mutations in the partial tree. This is a first step to study the asymptotic distribution of a natural estimator of DNA mutation rate for species with large families.

1. Introduction

1.1. **Motivations.** The Kingman coalescent, see [15, 16], allows to describe the genealogy of n individuals in a Wright-Fisher model, when the size of the whole population is very large and time is well rescaled. In what follows, we consider only neutral mutations and the infinite allele model, where each mutation gives a new allele. The Watterson estimator [22], based on the number of different alleles observed among n individuals alive today, $K^{(n)}$, allows to estimate the rate of mutation for the DNA, θ. This estimator is consistent and converges at rate $1/\sqrt{\log(n)}$.

Other models of population where one individual can produce a large number of children give rise to more general coalescent processes than the Kingman coalescent, where multiple collisions appear, see Sagitov [20] and Schweinsberg [21] (such models may be relevant for oysters and some fish species [7, 10]). In Birkner and al. [5] and in Schweinsberg [21] a natural family of one parameter coalescent processes arise to describe the genealogy of such populations: the Beta coalescent with parameter $\alpha \in (1, 2)$. Results from Beresticky and al. [2] give a consistent estimator, based on the observed number, $K^{(n)}$, of different alleles for the rate θ of mutation of DNA. This paper is a first step to study the convergence rate of this estimator or equivalently to the study the asymptotic distribution of $K^{(n)}$. Results are also known for the asymptotic distribution of $K^{(n)}$ for other coalescent processes, see Drmota and al. [9] and Möhle [17].

For the Beta coalescent, the asymptotic distribution of $K^{(n)}$ depends on θ but also on the parameter α. In particular, if the mutation rate of the DNA is known, the asymptotic distribution of $K^{(n)}$ allows to deduce an estimation and a confidence interval for α, which in a sense characterize the size of a typical family according to [21].

1.2. **The coalescent tree and mutation rate.** We consider at time $t = 0$ a number, $n \geq 1$ of individuals, and we look backward in time. Let $\mathcal{P}_n$ be the set of partitions of $\{1, \dots, n\}$. For $t \geq 0$, let $\Pi_t^{(n)}$ be an element of $\mathcal{P}_n$ such that each block of $\Pi_t^{(n)}$ corresponds to the initial individuals which have a common ancestor at time $-t$. We assume that if we consider b blocks, k of them merge into 1 at rate $\lambda_{b,k}$, independent of the current number of blocks. Using this property and the compatibility relation implied when one consider a larger number

Date: February 1, 2008.
2000 *Mathematics Subject Classification.* 60F05, 60G52, 60J70, 05C05.
Key words and phrases. Coalescent process, Beta-coalescent, stable process, Watterson estimator.

of initial individuals, Pitman [19], see also Sagitov [20] for a more biological approach, showed the transition rates are given by

$$\lambda_{b,k} = \int_{(0,1)} x^{k-2}(1-x)^{b-k}\Lambda(dx), \quad 2 \leq k \leq b,$$

for some finite measure Λ on $[0,1]$, and that $\Pi^{(n)}$ is the restriction of the so-called coalescent process defined on the set of partitions of $\mathbb{N}^*$. The Kingman coalescent correspond to the case where Λ is the Dirac mass at 0, see [15]. In particular, in the Kingman coalescent, only two blocks merge at a time. The Bolthausen-Sznitman [6] coalescent correspond to the case where Λ is the Lebesgue measure on $[0,1]$. The Beta-coalescent introduced in Birkner and al. [5] and in Schweinsberg [21], see also Bertoin and Le Gall [4] and Beresticky and al. [1] , corresponds to $\Lambda(dx) = C_0 x^{\alpha-1}(1-x)^{1-\alpha}\mathbf{1}_{(0,1)}(x)\, dx$ for some constant $C_0 > 0$.

Notice $\Pi^{(n)} = (\Pi_t^{(n)}, t \geq 0)$ is a Markov process starting at the trivial partition of $\{1, \ldots, n\}$ into n singletons. We denote by $R_t^{(n)}$ the number of blocks of $\Pi_t^{(n)}$, that is the number of common ancestors alive at time $-t$. In particular we have $R_0^{(n)} = n$. We shall omit the superscript (n) when there is no confusion. The process $R = (R_t, t \geq 0)$ is a continuous time Markov process taking values in $\mathbb{N}^*$. The number of possible choices of $\ell + 1$ blocks among k is $\binom{k}{\ell+1}$ (for $1 \leq \ell \leq k-1$) and each group of $\ell + 1$ blocks merge at rate $\lambda_{k,\ell+1}$. So the waiting time of R in state k is an exponential random variable with parameter

$$g_k = \sum_{\ell=1}^{k-1} \binom{k}{\ell+1} \lambda_{k,\ell+1} = \int_{(0,1)} \Big(1 - (1-x)^k - kx(1-x)^{k-1}\Big) \frac{\Lambda(dx)}{x^2} \tag{1}$$

and is distributed as E/g_k, where E is an exponential random variable with mean 1.

The apparition time of the most recent common ancestor (MRCA) is $T_n = \inf\{t > 0; R_t = 1\}$.

Let $Y = (Y_k, k \geq 1)$ be the different states of the process R. It is defined by $Y_0 = R_0$ and for $k \geq 1$, $Y_k = R_{S_k}$, where the sequence of jumping time $(S_k, k \geq 0)$ is defined inductively by $S_0 = 0$ and for $k \geq 1$, $S_k = \inf\{t > S_{k-1}; R_t \neq R_{S_{k-1}}\}$. We use the convention that $\inf \emptyset = +\infty$ and $Y_k = 1$ for $k \geq \tau_n$, where $\tau_n = \inf\{k; R_{S_k} = 1\}$ is the number of jumps of the process R until it reach the absorbing state 1. The number τ_n is the number of coalescences.

We shall write $Y^{(n)}$ instead of Y when it will be convenient to stress that Y starts at time 0 at point n. Notice Y is an $\mathbb{N}^*$-valued discrete time Markov chain, with probability transition

$$P(k, k-\ell) = \frac{\binom{k}{\ell+1}\lambda_{k,\ell+1}}{g_k}. \tag{2}$$

The sum of the lengths of all branches in the coalescent tree until the MRCA is distributed as

$$L^{(n)} = \sum_{k=0}^{\tau_n - 1} \frac{Y_k^{(n)}}{g_{Y_k^{(n)}}} E_k,$$

where $(E_k, k \geq 0)$ are independent exponential random variables with expectation 1.

In the infinite allele model, one assume that (neutral) mutations appear in the genealogy at random with rate θ. In particular by looking at the number $K^{(n)}$ of different alleles among n individuals, one get the number of mutations which occured in the genealogy of those individuals after the most recent common ancestor. In particular, conditionally on the length of the coalescent tree $L^{(n)}$, the number $K^{(n)}$ of mutations is distributed according to a Poisson

r.v. with parameter $\theta L^{(n)}$. Therefore, we have that $\dfrac{K^{(n)} - \theta L^{(n)}}{\sqrt{\theta L^{(n)}}}$ converges in distribution to a standard Gaussian r.v. (with mean 0 and variance 1). If the asymptotic distribution of $L^{(n)}$ is known, one can deduce the asymptotic distribution of $K^{(n)}$.

1.3. **Known results.**

1.3.1. *Kingman coalescence.* . For Kingman coalescence, a coalescence corresponds to the apparition of a common ancestor of only two individuals. In particular, we have for $0 \leq k \leq n-1$, $Y_k^{(n)} = n-k$. Thus we get $\tau_n = n-1$ as well as $g_{Y_k^{(n)}} = (n-k)(n-k-1)/2$. We also have $\dfrac{L^{(n)}}{2} = \sum_{k=0}^{n-2} \dfrac{1}{n-k-1} E_k = \sum_{k=1}^{n-1} \dfrac{1}{k} E_{n-k-1}$. The r.v. $L^{(n)}/2$ is distributed as the sum of independent exponential r.v. with parameter 1 to $n-1$, that is as the maximum on $n-1$ independent exponential r.v. with mean 1, see Feller [11] section I.6. An easy computation gives that $L^{(n)}/(2\log(n))$ converges in probability to 1 and that $\dfrac{L^{(n)}}{2} - \log(n)$ converges in distribution to the Gumbel distribution (with density $\mathrm{e}^{-x-\exp -x}$) when n goes to infinity. It is then easy to deduce that $\dfrac{K^{(n)} - \theta\mathbb{E}[L^{(n)}]}{\sqrt{\theta\mathbb{E}[L^{(n)}]}}$ converges in distribution to the standard Gaussian distribution. This provides the weak convergence and the asymptotic normality of the Watterson [22] estimator of θ: $\dfrac{K^{(n)}}{\mathbb{E}[L^{(n)}]} = \dfrac{K^{(n)}}{\sum_{k=1}^{n-1} \frac{1}{k}}$. See also the appendix in [9].

1.3.2. *Bolthausen-Sznitman coalescence.* In Drmota and al. [9], the authors consider the Bolthausen-Sznitman coalescence: Λ is the Lebesgue measure on $[0,1]$. In this case they prove that $\dfrac{1}{n}\log(n) L^{(n)}$ converges in probability to 1 and that $\dfrac{L^{(n)} - a_n}{b_n}$ converges in distribution to a stable r.v. Z with Laplace transform $\mathbb{E}[\mathrm{e}^{-\lambda Z}] = \mathrm{e}^{\lambda \log(\lambda)}$ for $\lambda > 0$, where

$$a_n = \frac{n}{\log(n)} + \frac{n \log(\log(n))}{\log(n)^2} \quad \text{and} \quad b_n = \frac{n}{\log(n)^2}.$$

It is then easy to deduce that $\dfrac{K^{(n)} - \theta a_n}{\theta b_n}$ converges to Z.

1.3.3. *The case $\int_{[0,1]} x^{-1}\Lambda(dx) < \infty$.* In Möhle [17], the author investigates the case where $x^{-1}\Lambda(dx)$ is a finite measure and consider directly the asymptotic distribution of $K^{(n)}$. In particular he gets that $K^{(n)}/n\theta$ converges in distribution to a non-negative r.v. Z uniquely determined by its moments: for $k \geq 1$,

$$\mathbb{E}[Z^k] = \frac{k!}{\prod_{i=1}^k \Phi(i)}, \quad \text{with} \quad \Phi(i) = \int_{[0,1]} (1-(1-x)^i) x^{-2} \Lambda(dx).$$

There is an equation in law for Z when Λ is a simple measure, that is when $\int_{[0,1]} x^{-2}\Lambda(dx) < \infty$.

1.3.4. *Beta coalescent.* The Beta-coalescent correspond to the case where Λ is the Beta$(2-\alpha,\alpha)$ distribution, with $\alpha \in (1,2)$: $\Lambda(dx) = \frac{1}{\Gamma(2-\alpha)\Gamma(\alpha)} x^{1-\alpha}(1-x)^{\alpha-1}dx$. The Kingman coalescent can be viewed as the asymptotic case $\alpha = 2$ and the Bolthausen-Sznitman coalescence as the asymptotic case $\alpha = 1$.

The first order asymptotic behavior of $L^{(n)}$ is given in [2], theorem 1.9: $n^{\alpha-2}L^{(n)}$ converges in probability to $\frac{\Gamma(\alpha)\alpha(\alpha-1)}{2-\alpha}$. We shall now investigate the asymptotic distribution of $L^{(n)}$.

1.4. **Main result.** In this paper we shall state a partial result concerning the asymptotic distribution of $L^{(n)}$. We shall only give the asymptotic distribution of the total length of the coalescent tree up to the $\lfloor nt \rfloor$-th coalescence:

$$L_t^{(n)} = \sum_{k=0}^{\lfloor nt \rfloor \wedge (\tau_n - 1)} \frac{Y_k^{(n)}}{g_{Y_k^{(n)}}} E_k, \tag{3}$$

where $\lfloor x \rfloor$ is the largest integer smaller or equal to x for $x \geq 0$.

We say $g = O(f)$, where f is a non-negative function and g a real valued function defined on a set E (mainly here $E = [0,1]$ or $E = \mathbb{N}^*$ or $E = \mathbb{N}^* \times [0,1]$), if there exists a finite constant $C > 0$ such that $|g(x)| \leq Cf(x)$ for all $x \in E$.

Let $\nu(dx) = x^{-2}\Lambda(dx)$ and $\rho(t) = \nu((t,1])$. We assume that $\rho(t) = C_0 t^{-\alpha} + O(t^{-\alpha+\zeta})$ for some $\alpha \in (1,2)$, $C_0 > 0$ and $\zeta > 1 - 1/\alpha$. This includes the Beta$(2-\alpha,\alpha)$ distribution for Λ. We have, see Lemma 2.2, that

$$g_n = C_0\Gamma(2-\alpha)n^\alpha + O(n^{\alpha-\min(\zeta,1)}).$$

Let $\gamma = \alpha - 1$. Let $V = (V_t, t \geq 0)$ be a α-stable Lévy process with no positive jumps (see chap. VII in [3]) with Laplace exponent $\psi(u) = u^\alpha/\gamma$: for all $u \geq 0$, $\mathbb{E}[\mathrm{e}^{-uV_t}] = \mathrm{e}^{tu^\alpha/\gamma}$.

We first give in Proposition 3.1 the asymptotic for the number of coalescences, τ_n:

$$n^{-\frac{1}{\alpha}}\left(n - \frac{\tau_n}{\gamma}\right) \xrightarrow[n\to\infty]{\text{(d)}} V_\gamma.$$

See also Gnedin and Yakubovich [12] and Iksanov and Möhle [13] for different proofs of this results under slightly different or stronger hypothesis. Then we give the asymptotics of $\hat{L}_t^{(n)}$ defined as $C_0\Gamma(2-\alpha)L_t^{(n)}$ but for the exponential r.v. E_k which are replaced by their mean that is 1 and for $g_{Y_k^{(n)}}$ which is replaced by its equivalent $C_0\Gamma(2-\alpha)\left(Y_k^{(n)}\right)^{2-\alpha}$:

$$\hat{L}_t^{(n)} = \sum_{k=0}^{\lfloor nt \rfloor \wedge (\tau_n - 1)} \left(Y_k^{(n)}\right)^{1-\alpha}. \tag{4}$$

For $t \in [0,\gamma]$, we set

$$v(t) = \int_0^t \left(1 - \frac{r}{\gamma}\right)^{-\gamma} dr.$$

Theorem 5.1 gives that the following convergence in distribution holds for all $t \in (0,\gamma)$

$$n^{-1+\alpha-1/\alpha}(\hat{L}_t^{(n)} - n^{2-\alpha}v(t)) \xrightarrow[n\to\infty]{\text{(d)}} (\alpha-1)\int_0^t dr\ (1-\frac{r}{\gamma})^{-\alpha} V_r. \tag{5}$$

Then we deduce our main result, Theorem 6.1. Let $\alpha \in (1, \frac{1+\sqrt{5}}{2})$. Then for all $t \in (0, \gamma)$, we have the following convergence in distribution

$$n^{-1+\alpha-1/\alpha}\left(L_t^{(n)} - n^{2-\alpha}\frac{v(t)}{C_0\Gamma(2-\alpha)}\right) \xrightarrow[n\to\infty]{\text{(d)}} \frac{\alpha-1}{C_0\Gamma(2-\alpha)}\int_0^t dr\ (1-\frac{r}{\gamma})^{-\alpha}V_r. \tag{6}$$

We also have that $n^{\alpha-2}L_t^{(n)}$ converges in probability to $\frac{v(t)}{C_0\Gamma(2-\alpha)}$ for $\alpha \in (1,2)$. For $t = \gamma$, intuitively we have $L_\gamma^{(n)}$ close to $L^{(n)}$ as τ_n is close to n/γ. In particular, one expects that $n^{\alpha-2}L^{(n)}$ converges in probability to $\frac{v(\gamma)}{C_0\Gamma(2-\alpha)}$. For the Beta-coalescent, $\Lambda(dx) = \frac{1}{\Gamma(2-\alpha)\Gamma(\alpha)}\, x^{1-\alpha}(1-x)^{\alpha-1}dx$, we have $C_0 = 1/\alpha\Gamma(2-\alpha)\Gamma(\alpha)$ and indeed, theorem 1.9 in [2] gives that $n^{\alpha-2}L^{(n)}$ converges in probability to $\frac{\Gamma(\alpha)\alpha(\alpha-1)}{2-\alpha} = \frac{v(\gamma)}{C_0\Gamma(2-\alpha)}$. Notice theorem 1.9 in [2] is stated for more general coalescents than the Beta-coalescent.

In Corollary 6.2, we give the asymptotic distribution of the number $K_t^{(n)}$ of mutations on the coalescent tree up to the $\lfloor nt \rfloor$-th coalescent for $\alpha \in (1,2)$. In particular, for $\alpha > \frac{1+\sqrt{5}}{2}$, the approximations of the exponential r.v. by their mean are more important than the fluctuations of $\hat{L}^{(n)}$, and the asymptotic distribution is gaussian.

1.5. **Organization of the paper.** In Section 2 we give estimates (distribution, Laplace transform) for the number of collisions in the first coalescence in a population of n individuals. We prove the asymptotic distribution of the number of collisions, τ_n, in Section 3, as well as an invariance principle for the coalescent process $Y^{(n)}$, see Corollary 3.5. In Section 4, we give error bounds on the approximation of $L_t^{(n)}$ by $\hat{L}_t^{(n)}/C_0\Gamma(2-\alpha)$. Section 5 is devoted to the asymptotic distribution of $\hat{L}_t^{(n)}$. Eventually, our main result, Theorem 6.1, on the asymptotic distribution of $L_t^{(n)}$, and Corollary 6.2, on the asymptotic distribution of the number of mutations $K_t^{(n)}$, and their proofs are given in Section 6.

In what follows, c is a non important constant which value may vary from line to line.

2. Law of the first jump

Let Y be a discrete time Markov chain on $\mathbb{N}^*$ with transition kernel P given by (2) and started at $Y_0 = n$. Let $X_k^{(n)} = Y_{k-1} - Y_k$ for $k \geq 1$. We give some estimates on the moment of $X_1^{(n)}$ and its Laplace transform.

For $n \geq 1$, $x \in (0,1)$, let $B_{n,x}$ be a binomial r.v. with parameter (n,x). Recall that for $1 \leq k \leq n$, we have

$$\mathbb{P}(B_{n,x} \geq k) = \frac{n!}{(k-1)!(n-k)!}\int_0^x t^{k-1}(1-t)^{n-k}\, dt. \tag{7}$$

Recall that $\nu(dx) = x^{-2}\Lambda(dx)$ and $\rho(t) = \nu((t,1])$. Use the first equality in (1) and (7) to get

$$
\begin{aligned}
g_n &= \int_0^1 \sum_{k=2}^n \binom{n}{k} x^k (1-x)^{n-k} \nu(dx) \\
&= \int_0^1 \mathbb{P}(B_{n,x} \geq 2)\nu(dx) \\
&= n(n-1)\int_0^1 (1-t)^{n-2} t\rho(t)\, dt. \qquad (8)
\end{aligned}
$$

Notice also that $\mathbb{P}(X_1^{(n)} = k) = P(n, n-k) = \dfrac{1}{g_n}\displaystyle\int_0^1 \mathbb{P}(B_{n,x} = k+1)\nu(dx)$ and thus

$$
\mathbb{P}(X_1^{(n)} \geq k) = \frac{\int_0^1 \mathbb{P}(B_{n,x} \geq k+1)\nu(dx)}{g_n} = \frac{(n-2)!}{k!(n-k-1)!} \frac{\int_0^1 (1-t)^{n-k-1} t^k \rho(t)\, dt}{\int_0^1 (1-t)^{n-2} t\rho(t)\, dt}. \qquad (9)
$$

Let $\alpha \in (1,2)$ and $\gamma = \alpha - 1$.

We say $g = o(f)$, where f is a non-negative function and g a real valued function defined on $(0,1]$, if for any $\varepsilon > 0$, there exists $x_0 > 0$ s.t. $|g(x)| \leq \varepsilon f(x)$ for all $x \in (0, x_0]$.

Lemma 2.1. *Assume that $\rho(t) = C_0 t^{-\alpha} + o(t^{-\alpha})$. Then $(X_1^{(n)}, n \geq 2)$ converges in distribution to the r.v. X such that for all $k \geq 1$,*

$$
\mathbb{P}(X \geq k) = \frac{1}{\Gamma(2-\alpha)} \frac{\Gamma(k+1-\alpha)}{k!}.
$$

We have $\mathbb{E}[X] = 1/\gamma$, $\mathbb{E}[X^2] = +\infty$ and its Laplace transform ϕ is given by: for $u \geq 0$,

$$
\phi(u) = \mathbb{E}[\mathrm{e}^{-uX}] = 1 + \frac{\mathrm{e}^u - 1}{\alpha - 1}\left[(1-\mathrm{e}^{-u})^{\alpha-1} - 1\right].
$$

We shall use repeatedly the identity of the beta distribution: for $a > 0$ and $b > 0$, we have

$$
\int_0^1 t^{a-1}(1-t)^{b-1} dt = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}. \qquad (10)
$$

Proof. The condition $\rho(t) = C_0 t^{-\alpha} + o(t^{-\alpha})$ implies that for fixed $k \geq 1$, as n goes to infinity, we have

$$
\int_0^1 (1-t)^{n-k-1} t^k \rho(t)\, dt = \frac{\Gamma(k+1-\alpha)\Gamma(n-k)}{\Gamma(n+1-\alpha)}\left(C_0 + o(1)\right).
$$

Therefore, we get that

$$
\begin{aligned}
\lim_{n\to\infty} \mathbb{P}(X_1^{(n)} \geq k) &= \lim_{n\to\infty} \frac{(n-2)!}{k!(n-k-1)!} \frac{\int_0^1 (1-t)^{n-k-1} t^k \rho(t)\, dt}{\int_0^1 (1-t)^{n-2} t\rho(t)\, dt} \\
&= \lim_{n\to\infty} \frac{(n-2)!}{k!(n-k-1)!} \frac{\Gamma(k+1-\alpha)\Gamma(n-k)}{\Gamma(n+1-\alpha)} \frac{\Gamma(n+1-\alpha)}{\Gamma(2-\alpha)\Gamma(n-1)} \\
&= \frac{1}{\Gamma(2-\alpha)} \frac{\Gamma(k+1-\alpha)}{k!}.
\end{aligned}
$$

This ends the first part of the Lemma. Notice that

$$
\mathbb{P}(X \geq k) = \frac{1}{\Gamma(\alpha)\Gamma(2-\alpha)} \int_0^1 t^{k-\alpha}(1-t)^{\alpha-1} dt
$$

and as $\mathbb{P}(X = k) = \mathbb{P}(X \geq k) - \mathbb{P}(X \geq k+1)$, we get

$$\mathbb{P}(X = k) = \frac{1}{\Gamma(\alpha)\Gamma(2-\alpha)} \int_0^1 t^{k-\alpha}(1-t)^{\alpha} dt = \frac{\alpha}{\Gamma(2-\alpha)} \frac{\Gamma(k+1-\alpha)}{(k+1)!}. \tag{11}$$

We have

$$\begin{aligned}
\mathbb{E}[X] = \sum_{k\geq 1} \mathbb{P}(X \geq k) &= \frac{1}{\Gamma(\alpha)\Gamma(2-\alpha)} \int_0^1 \sum_{k\geq 1} t^{k-\alpha}(1-t)^{\alpha-1} dt \\
&= \frac{1}{\Gamma(\alpha)\Gamma(2-\alpha)} \int_0^1 t^{1-\alpha}(1-t)^{\alpha-2} dt \\
&= \frac{1}{\Gamma(\alpha)\Gamma(2-\alpha)} \frac{\Gamma(2-\alpha)\Gamma(\alpha-1)}{\Gamma(1)} \\
&= \frac{1}{\alpha-1}.
\end{aligned}$$

The asymptotic expansion

$$\Gamma(z) = \sqrt{2\pi} z^{z-1/2} \mathrm{e}^{-z} \left(1 + \frac{1}{12z} + o\left(\frac{1}{z}\right)\right) \tag{12}$$

implies $\mathbb{P}(X = k) \sim_{+\infty} \dfrac{\alpha}{\Gamma(2-\alpha)} k^{-\alpha-1}$. Therefore we have $\mathbb{E}[X^2] = +\infty$. We compute the Laplace transform of X. Let $u \geq 0$, we have

$$\begin{aligned}
\phi(u) = \mathbb{E}[\mathrm{e}^{-uX}] &= \frac{\alpha}{\Gamma(2-\alpha)} \sum_{k\geq 1} \frac{1}{(k+1)!} \mathrm{e}^{-ku} \int_0^\infty x^{k-\alpha} \mathrm{e}^{-x} \, dx \\
&= \frac{\alpha \, \mathrm{e}^{u}}{\Gamma(2-\alpha)} \int_0^\infty \sum_{k\geq 2} \frac{1}{k!} \mathrm{e}^{-ku} x^{k-1-\alpha} \mathrm{e}^{-x} \, dx \\
&= \frac{\alpha \, \mathrm{e}^{u}}{\Gamma(2-\alpha)} \int_0^\infty x^{-1-\alpha} \mathrm{e}^{-x} (\mathrm{e}^{x \, \mathrm{e}^{-u}} - x \, \mathrm{e}^{-u} - 1) \, dx \\
&= 1 + \frac{\mathrm{e}^{u} - 1}{\alpha - 1} \left[(1 - \mathrm{e}^{-u})^{\alpha-1} - 1\right],
\end{aligned}$$

where we used (11) with $\Gamma(k+1-\alpha) = \displaystyle\int_0^\infty x^{k-\alpha} \mathrm{e}^{-x} \, dx$ for the first equality and two integrations by parts for the last. □

We give bounds on g_n.

Lemma 2.2. *Assume that $\rho(t) = C_0 t^{-\alpha} + O(t^{-\alpha+\zeta})$ for some $C_0 > 0$ and $\zeta > 0$. Then we have, for $n \geq 2$,*

$$g_n = C_0 \Gamma(2-\alpha) n^{\alpha} + O(n^{\alpha - \min(\zeta,1)}). \tag{13}$$

Proof. Notice that

$$g_n = n(n-1) \int_0^1 (1-t)^{n-2} t \left(C_0 t^{-\alpha} + O(t^{-\alpha+\zeta})\right) \, dt = C_0 n(n-1) \frac{\Gamma(2-\alpha)\Gamma(n-1)}{\Gamma(n+1-\alpha)} + h_n,$$

where $h_n = n(n-1)\int_0^1 (1-t)^{n-2} t^{-\alpha+\zeta+1} O(1)\ dt$. In particular, using (12), we have for $n \geq 2$

$$|h_n| \leq cn(n-1)\int_0^1 (1-t)^{n-2} t^{-\alpha+\zeta+1} = cn(n-1)\frac{\Gamma(2-\alpha+\zeta)\Gamma(n-1)}{\Gamma(n+1-\alpha+\zeta)} \leq cn^{\alpha-\zeta}.$$

Using (12) again, we get that $\Gamma(n-1)/\Gamma(n+1-\alpha) = n^{\alpha-2} + O(n^{\alpha-3})$. This implies that

$$g_n = C_0\Gamma(2-\alpha)n^\alpha + O(n^{\max(\alpha-1,\alpha-\zeta)}).$$

□

We give an expansion of the first moment of $X_1^{(n)}$.

Lemma 2.3. *Assume that $\rho(t) = C_0 t^{-\alpha} + O(t^{-\alpha+\zeta})$ for some $C_0 > 0$ and $\zeta > 0$. Let $\varepsilon_0 > 0$. We set*

$$\varphi_n = \begin{cases} n^{-\zeta} & \text{if} \quad \zeta < \alpha - 1, \\ n^{1-\alpha+\varepsilon_0} & \text{if} \quad \zeta = \alpha - 1, \\ n^{1-\alpha} & \text{if} \quad \zeta > \alpha - 1. \end{cases} \tag{14}$$

There exists a constant C_{15} s.t. for all $n \geq 2$, we have

$$\left| \mathbb{E}[X_1^{(n)}] - \frac{1}{\gamma} \right| \leq C_{15}\varphi_n. \tag{15}$$

Proof. We have

$$\begin{aligned} \mathbb{E}[X_1^{(n)}] = \sum_{k\geq 1} \mathbb{P}(X_1^{(n)} \geq k) &= \frac{\int_0^1 \sum_{k\geq 1} \mathbb{P}(B_{n,x} \geq k+1)\nu(dx)}{g_n} \\ &= \frac{\int_0^1 (\mathbb{E}[B_{n,x}] - \mathbb{P}(B_{n,x} \geq 1))\nu(dx)}{g_n} \qquad (16) \\ &= \frac{\int_0^1 nx\nu(dx) - \int_0^1 (1-(1-x)^n)\nu(dx)}{g_n} \\ &= \frac{n\int_0^1 [1-(1-t)^{n-1}]\rho(t)\ dt}{g_n} \qquad (17) \\ &= \frac{\int_0^1 (1-t)^{n-2}\left(\int_t^1 \rho(r)\ dr\right)\ dt}{\int_0^1 (1-t)^{n-2} t\rho(t)\ dt}, \end{aligned}$$

using (9) for the first equality and (8) for the last. Notice that

$$\begin{aligned} \int_t^1 \rho(r)\ dr &= \frac{1}{\gamma} t\rho(t) + O(1) + \int_t^1 O(r^{-\alpha+\zeta})\ dr + O(t^{-\alpha+\zeta+1}) \\ &= \frac{1}{\gamma} t\rho(t) + O(t^{\min(-\alpha+\zeta+1,0)}) + O(|\log(t)|)\mathbf{1}_{\{\alpha-\zeta=1\}} \\ &= \frac{1}{\gamma} t\rho(t) + O(t^{\min(-\alpha+\zeta+1,0)}) + O(t^{-\varepsilon_0})\mathbf{1}_{\{\alpha-\zeta=1\}}. \end{aligned}$$

This implies that

$$\mathbb{E}[X_1^{(n)}] = \frac{1}{\gamma} + \frac{n(n-1)}{g_n}\int_0^1 (1-t)^{n-2}\left(O(t^{\min(-\alpha+\zeta+1,0)}) + O(t^{-\varepsilon_0})\mathbf{1}_{\{\alpha-\zeta=1\}}\right)\ dt.$$

Using (10), (12) and Lemma 2.2, we get

$$\begin{aligned}\left|\mathbb{E}[X_1^{(n)}] - \frac{1}{\gamma}\right| &\leq c\frac{n(n-1)}{g_n}\int_0^1 (1-t)^{n-2}\left(t^{\min(-\alpha+\zeta+1,0)} + t^{-\varepsilon_0}\mathbf{1}_{\{\alpha-\zeta=1\}}\right)\ dt \\ &\leq cn^{2-\alpha}(n^{-1-\min(-\alpha+\zeta+1,0)} + n^{-1+\varepsilon_0}\mathbf{1}_{\{\alpha-\zeta=1\}}) \\ &\leq c\varphi_n.\end{aligned}$$

□

We give an upper bound for the second moment of $X_1^{(n)}$.

Lemma 2.4. *Assume that $\rho(t) = O(t^{-\alpha})$. Then there exists a constant C_{18} s.t. for all $n \geq 2$, we have*

$$\mathbb{E}\left[\left(X_1^{(n)}\right)^2\right] \leq C_{18}\frac{n^2}{g_n}. \tag{18}$$

Proof. Using the identity $\mathbb{E}[Y^2] = \sum_{k\geq 1}(2k-1)\mathbb{P}(Y \geq k)$ for $\mathbb{N}$-valued random variables, we get

$$\begin{aligned}\mathbb{E}\left[\left(X_1^{(n)}\right)^2\right] &= \frac{\int_0^1 \sum_{k\geq 1}(2k-1)\mathbb{P}(B_{n,x} \geq k+1)\nu(dx)}{g_n} \\ &= \frac{\int_0^1 \left(\sum_{k\geq 1}(2(k+1)-1)\mathbb{P}(B_{n,x} \geq k+1) - 2\sum_{k\geq 1}\mathbb{P}(B_{n,x} \geq k+1)\right)\nu(dx)}{g_n} \\ &= \frac{\int_0^1 \left(\mathbb{E}[B_{n,x}^2] - 2\mathbb{E}[B_{n,x}] + \mathbb{P}(B_{n,x} \geq 1)\right)\nu(dx)}{g_n} \\ &= \frac{\int_0^1 \left(\mathbb{E}[B_{n,x}^2] - \mathbb{E}[B_{n,x}]\right)\nu(dx)}{g_n} - \mathbb{E}[X_1^{(n)}] \\ &= \frac{\int_0^1 n(n-1)x^2\nu(dx)}{g_n} - \mathbb{E}[X_1^{(n)}] \\ &= 2n(n-1)\frac{\int_0^1 t\rho(t)\ dt}{g_n} - \mathbb{E}[X_1^{(n)}],\end{aligned}$$

where we have used (16) for the fourth equality. Use $\int_0^1 t\rho(t)\ dt < \infty$ and $\mathbb{E}[X_1^{(n)}] \geq 0$ to conclude. □

We consider ϕ_n the Laplace transform of $X_1^{(n)}$: for $u \geq 0$, $\phi_n(u) = \mathbb{E}[\mathrm{e}^{-uX_1^{(n)}}]$.

Lemma 2.5. *Assume that $\rho(t) = C_0t^{-\alpha} + O(t^{-\alpha+\zeta})$ for some $C_0 > 0$ and $\zeta > 0$. Let $\varepsilon_0 > 0$. Recall φ_n given by (14). Then we have, for $n \geq 2$,*

$$\phi_n(u) = 1 - \frac{u}{\gamma} + \frac{u^\alpha}{\gamma} + R(n,u), \tag{19}$$

where $R(n,u) = \left(u\varphi_n + u^2\right)h(n,u)$ with $\sup_{u\in[0,K],n\geq 2}|h(n,u)| < \infty$.

Proof. We have

$$\begin{aligned}
&\phi_n(u)\\
&\quad= \mathbb{E}\left[\mathrm{e}^{-uX_1^{(n)}}\right] = \sum_{k=1}^{n-1} \mathrm{e}^{-uk}\,\mathbb{P}(X_1^{(n)}=k)\\
&\quad= \sum_{k=1}^{n-1} \mathrm{e}^{-uk}\,\mathbb{P}(X_1^{(n)}\geq k) - \sum_{k=2}^{n} \mathrm{e}^{-u(k-1)}\,\mathbb{P}(X_1^{(n)}\geq k)\\
&\quad= \mathrm{e}^{-u} + \sum_{k=2}^{n-1} \mathrm{e}^{-uk}(1-\mathrm{e}^{u})\mathbb{P}(X_1^{(n)}\geq k)\\
&\quad= \mathrm{e}^{-u} + (1-\mathrm{e}^{u})\sum_{k=2}^{n-1} \frac{\mathrm{e}^{-uk}}{g_n}\int_0^1 \frac{n!}{k!(n-k-1)!} t^k(1-t)^{n-k-1}\rho(t)\,dt\\
&\quad= \mathrm{e}^{-u} + (1-\mathrm{e}^{u})\frac{n}{g_n}\int_0^1 \left[(1-t(1-\mathrm{e}^{-u}))^{n-1} - (1-t)^{n-1} - (n-1)\,\mathrm{e}^{-u}\,t(1-t)^{n-2}\right]\rho(t)dt\\
&\quad= 1 + (1-\mathrm{e}^{u})\frac{n}{g_n}\int_0^1 \left[(1-t(1-\mathrm{e}^{-u}))^{n-1} - (1-t)^{n-1}\right]\rho(t)\,dt,
\end{aligned}$$

where we used (8) for the last equality. Using (17), this implies

$$\phi_n(u) = 1 + (1-\mathrm{e}^{u})\frac{n}{g_n}A + (1-\mathrm{e}^{u})\mathbb{E}[X_1^{(n)}]. \tag{20}$$

with $A = \int_0^1 \left[(1-t(1-\mathrm{e}^{-u}))^{n-1} - 1\right]\rho(t)\,dt$.

Thanks to Lemma 2.3, we have that

$$(1-\mathrm{e}^{u})\mathbb{E}[X_1^{(n)}] = -\frac{u}{\gamma} + \left(u^2 + u\varphi_n\right) h_1(n,u), \tag{21}$$

where $\sup_{u\in[0,K],n\geq 2} |h_1(n,u)| < \infty$.

To compute A, we set $a = (1-\mathrm{e}^{-u})$ and $f(t) = t^{-\max(\alpha-1-\zeta,0)} + t^{-\varepsilon_0}\mathbf{1}_{\{\alpha-\zeta=1\}}$. An integration by part gives

$$\begin{aligned}
A &= -a(n-1)\int_0^1 (1-at)^{n-2}\left(\int_t^1 \rho(r)\,dr\right)\,dt\\
&= -a(n-1)C_0\int_0^1 (1-at)^{n-2}\left(\frac{t^{1-\alpha}}{\gamma} + O(f(t))\right)\,dt\\
&= -A_1 + A_2,
\end{aligned}$$

with $A_1 = \dfrac{a(n-1)}{\gamma} C_0 \displaystyle\int_0^1 (1-at)^{n-2} t^{1-\alpha}\, dt$ and $A_2 = a(n-1) \displaystyle\int_0^1 (1-at)^{n-2} O(f(t))\, dt$. We have

$$\begin{aligned}
A_1 &= \frac{a^{\alpha-1}(n-1)}{\gamma} C_0 \int_0^a (1-t)^{n-2} t^{1-\alpha}\, dt \\
&= \frac{a^{\alpha-1}(n-1)}{\gamma} C_0 \int_0^1 (1-t)^{n-2} t^{1-\alpha}\, dt - \frac{a^{\alpha-1}(n-1)}{\gamma} C_0 \int_a^1 (1-t)^{n-2} t^{1-\alpha}\, dt \\
&= \frac{a^{\alpha-1}(n-1)}{\gamma} C_0 \frac{\Gamma(n-1)\Gamma(2-\alpha)}{\Gamma(n+1-\alpha)} - \frac{a^{\alpha-1}(n-1)}{\gamma} C_0 \int_a^1 (1-t)^{n-2} t^{1-\alpha}\, dt
\end{aligned}$$

Since $a \geq 0$, we have for $u \in [0, K]$ and $n \geq 2$

$$0 \leq \frac{a^{\alpha-1}(n-1)}{\gamma} \int_a^1 (1-t)^{n-2} t^{1-\alpha}\, dt \leq \frac{(n-1)}{\gamma} \int_a^1 (1-t)^{n-2}\, dt \leq \frac{1}{\gamma}\,.$$

Using (12) and Lemma 2.2, we get $|A_1 - \dfrac{a^{\alpha-1}}{\gamma}\dfrac{g_n}{n}| \leq c(1 + n^{\alpha-1-\min(\zeta,1)}) \leq c n^{\max(\alpha-1-\zeta,0)}$, where c does not depend on n and $u \geq 0$. We also have, using (10) and (12)

$$|A_2| \leq ca(n-1) \int_0^1 (1-at)^{n-2} f(t)\, dt \leq c(n^{\max(\alpha-1-\zeta,0)} + n^{\varepsilon_0} \mathbf{1}_{\{\alpha-\zeta=1\}}).$$

We deduce, using Lemma 2.2 twice, that

$$|A + \frac{a^{\alpha-1}}{\gamma}\frac{g_n}{n}| \leq c(n^{\max(\alpha-1-\zeta,0)} + n^{\varepsilon_0} \mathbf{1}_{\{\alpha-\zeta=1\}}) \leq c\frac{g_n}{n}\varphi_n.$$

We deduce that

$$(1-\mathrm{e}^u)\frac{n}{g_n} A = (1-\mathrm{e}^u)\left(-\frac{(1-\mathrm{e}^{-u})^{\alpha-1}}{\gamma} + \varphi_n O(1)\right) = \frac{u^\alpha}{\gamma} + \left(u^{\alpha+1} + u\varphi_n\right) h_2(n,u), \tag{22}$$

where $\sup_{u\in[0,K], n\geq 2} |h_2(n,u)| < \infty$. Then use the expression of ϕ_n given by (20) as well as (21) and (22) to end the proof. □

3. Asymptotics for the number of jumps

Let $\alpha \in (1,2)$. We assume that $\rho(t) = C_0 t^{-\alpha} + O(t^{-\alpha+\zeta})$ for some $C_0 > 0$ and $\zeta > 1 - 1/\alpha$.

Let $V = (V_t, t \geq 0)$ be a α-stable Lévy process with no positive jumps (see chap. VII in [3]) with Laplace exponent $\psi(u) = u^\alpha/\gamma$: for all $u \geq 0$, $\mathbb{E}[\mathrm{e}^{-uV_t}] = \mathrm{e}^{tu^\alpha/\gamma}$.

Lemma 2.1 implies that $(X_1^{(n)}, \dots, X_k^{(n)})$ converges in distribution to $(X_1, \dots, X_k)$ where $(X_k, k \geq 1)$ is a sequence of independent random variables distributed as X. Using Lemma 2.1 and (12), we get that $\mathbb{P}(X \geq k) \sim_{+\infty} \frac{1}{\Gamma(2-\alpha)} k^{-\alpha}$. Hence Proposition 9.39 in [8] implies that the law of X is in the domain of attraction of the α-stable distribution. We set $W_t^{(n)} = n^{-1/\alpha} \displaystyle\sum_{k=1}^{\lfloor nt \rfloor} (X_k - \frac{1}{\gamma})$ for $t \in [0, \gamma]$. An easy calculation using the Laplace transform of X shows that for fixed t the sequence $W_t^{(n)}$ converges in distribution to V_t. Then using Theorem 16.14 in [14], we get that the process $(W_t^{(n)}, t \in [0,\gamma])$ converges in distribution to $V = (V_t, t \in [0,\gamma])$. We shall give in Corollary 3.5 a similar result with X_k replaced by $X_k^{(n)}$.

We first give a proof of the convergence of τ_n, see also [12] and [9] for a different proof. We will use that $\sum_{i=1}^{\tau_n}(X_i^{(n)} - \frac{1}{\gamma}) = n - 1 - \frac{\tau_n}{\gamma}$.

Proposition 3.1. *We assume that $\zeta > 1 - 1/\alpha$. We have the following convergence is distribution*

$$n^{-\frac{1}{\alpha}}\left(n - \frac{\tau_n}{\gamma}\right) \xrightarrow[n\to\infty]{(d)} V_\gamma.$$

Proof. Using [18], it is enough to prove that $\lim_{n\to\infty} \mathbb{E}[\mathrm{e}^{-un^{-\frac{1}{\alpha}}\left(n-\frac{\tau_n}{\gamma}\right)}] = \mathrm{e}^{u^\alpha}$ for all $u \geq 0$. Let $\mathcal{Y} = (\mathcal{Y}_k, k \geq 0)$ be the filtration generated by Y. Notice τ_n is an $\mathcal{Y}$-stopping time. For fixed n, and for any $v \geq 0$, the process $(M_{v,k}, k \geq 0)$ defined by

$$M_{v,k} = \prod_{i=1}^{k}\left(\exp -vX_i^{(n)} - \log \phi_{Y_{i-1}^{(n)}}(v)\right)$$

is a bounded martingale w.r.t. the filtration $\mathcal{Y}$. Notice that $\mathbb{E}[M_{v,k}] = 1$. As $X_i = 0$ for $i > \tau_n$, we also have

$$M_{v,k} = \prod_{i=1}^{k\wedge\tau_n}\left(\exp -vX_i^{(n)} - \log \phi_{Y_{i-1}^{(n)}}(v)\right). \tag{23}$$

Let $u \geq 0$ and consider a non-negative sequence $(a_n, n \geq 1)$ which converges to 0. Using (19), we get that :

$$M_{ua_n,k} = \exp\left(-ua_n\sum_{i=1}^{k\wedge\tau_n} X_i^{(n)} - \sum_{i=1}^{k\wedge\tau_n}\left(-\frac{ua_n}{\gamma} + \frac{u^\alpha a_n^\alpha}{\gamma} + R(Y_{i-1}^{(n)}, ua_n)\right)\right).$$

In particular, we have

$$M_{ua_n,\tau_n} = \exp\left(-ua_n(n - 1 - \frac{\tau_n}{\gamma}) - \frac{u^\alpha \tau_n a_n^\alpha}{\gamma} - \sum_{i=1}^{\tau_n} R(Y_{i-1}^{(n)}, ua_n)\right). \tag{24}$$

We first give an upper bound for $\sum_{i=1}^{\tau_n} R(Y_{i-1}^{(n)}, ua_n)$.

Lemma 3.2. *We assume that $\zeta > 1 - 1/\alpha$. Let $K > 0$. Let $\eta \geq \frac{1}{\alpha}$. There exist $\varepsilon_1 > 0$ and $C_{25}(K)$ a finite constant such that for all $n \geq 1$ and $u \in [0, K]$, a.s. with $a_n = n^{-\eta}$,*

$$\sum_{i=1}^{\tau_n}\left|R(Y_{i-1}^{(n)}, ua_n)\right| \leq C_{25}(K)n^{-\varepsilon_1}. \tag{25}$$

Proof. Notice that $\tau_n \leq n-1$. We have seen in Lemma 2.5 that $R(n, u) = \left(u\varphi_n + u^2\right) h(n, u)$ with $\bar{h}(K) = \sup_{u\in[0,K],n\geq 2} |h(n, u)| < \infty$ and φ_n given by (14). We have $2 - \alpha - \dfrac{1}{\alpha} = -\alpha(1 - 1/\alpha)^2 < 0$. As $\varepsilon_0 > 0$ is arbitrary in (14), we can take ε_0 small enough so that $1 - \alpha + \varepsilon_0 < 0$ and $2 - \alpha + \varepsilon_0 - 1/\alpha < 0$. We have

$$a_n\sum_{i=1}^{\tau_n}\varphi_{Y_{i-1}^{(n)}} \leq n^{-1/\alpha}\sum_{j=1}^{n}\varphi_j \leq c\begin{cases} n^{1-\zeta-\frac{1}{\alpha}} & \text{if} \quad \zeta < \alpha - 1, \\ n^{2-\alpha+\varepsilon_0-\frac{1}{\alpha}} & \text{if} \quad \zeta = \alpha - 1, \\ n^{2-\alpha-\frac{1}{\alpha}} & \text{if} \quad \zeta > \alpha - 1. \end{cases}$$

For $\varepsilon_1 > 0$ less than the two positive quantities $-1+\zeta+\dfrac{1}{\alpha}$ and $-2+\alpha-\varepsilon_0+\dfrac{1}{\alpha}$, we have $a_n \sum_{i=1}^{\tau_n} \varphi_{Y_{i-1}^{(n)}} \leq cn^{-\varepsilon_1}$. We deduce that, for $u \in [0,K]$,

$$\begin{aligned}
\sum_{i=1}^{\tau_n} \left| R(Y_{i-1}^{(n)}, ua_n) \right| &\leq \bar{h}(K) \sum_{i=1}^{\tau_n} \left(\varphi_{Y_{i-1}^{(n)}} ua_n + (ua_n)^2 \right) \\
&\leq \bar{h}(K) \sum_{j=1}^{n} \left(\varphi_j K a_n + (Ka_n)^2 \right) \\
&\leq c\bar{h}(K)(Kn^{-\varepsilon_1} + K^2 n^{1-\frac{2}{\alpha}}),
\end{aligned}$$

for some constant c independent of n, u and K. Taking $\varepsilon_1 > 0$ small enough so that $\varepsilon_1 < \dfrac{2}{\alpha} - 1$, we then get (25). □

Next we prove the following Lemma.

Lemma 3.3. *We assume that* $\zeta > 1 - 1/\alpha$. *Let* $\varepsilon > 0$. *The sequence* $(n^{-(1/\alpha)-\varepsilon}(n-1-\frac{\tau_n}{\gamma}), n \geq 1)$ *converges in probability to* 0.

Proof. We set $a_n = n^{-\frac{1}{\alpha}-\varepsilon}$. Notice that

$$\mathrm{e}^{-ua_n(n-1-\frac{\tau_n}{\gamma})} = M_{ua_n,\tau_n}\, \mathrm{e}^{\frac{u^\alpha \tau_n a_n}{\gamma} + \sum_{i=1}^{\tau_n} R(Y_{i-1}^{(n)}, ua_n)} .$$

As $\tau_n \leq n-1$, we have $0 \leq \tau_n a_n^\alpha \leq n^{-\alpha\varepsilon}$. Using (25), we get for $u \geq 0$

$$\mathbb{E}[M_{ua_n,\tau_n}]\, \mathrm{e}^{-C} 25^{(u)n^{-\varepsilon_1}} \leq \mathbb{E}[\mathrm{e}^{-ua_n(n-1-\frac{\tau_n}{\gamma})}] \leq \mathbb{E}[M_{ua_n,\tau_n}]\, \mathrm{e}^{C} 25^{(u)n^{-\varepsilon_1} + \frac{u^\alpha n^{-\alpha\varepsilon}}{\gamma}} .$$

As τ_n is bounded, the stopping time theorem gives $\mathbb{E}[M_{ua_n,\tau_n}] = 1$. We deduce that, for all $u \geq 0$, $\lim_{n\to\infty} \mathbb{E}[\mathrm{e}^{-ua_n(n-1-\frac{\tau_n}{\gamma})}] = 1$. Using [18], we get the convergence in law of $a_n(n-1-\frac{\tau_n}{\gamma})$ to 0, and then in probability as the limit is constant. □

Let $a_n = n^{-\frac{1}{\alpha}}$ and $u \geq 0$. We have

$$\begin{aligned}
(26)\quad \mathbb{E}\left[\mathrm{e}^{-ua_n(n-1-\frac{\tau_n}{\gamma})}\right] \\
&= \mathbb{E}\left[\mathrm{e}^{-ua_n(n-1-\frac{\tau_n}{\gamma})} \left(1 - \mathrm{e}^{-u^\alpha a_n^\alpha (\frac{\tau_n}{\gamma} - n)}\right)\right] + \mathbb{E}\left[\mathrm{e}^{-ua_n(n-1-\frac{\tau_n}{\gamma})}\, \mathrm{e}^{-u^\alpha a_n^\alpha (\frac{\tau_n}{\gamma} - n)}\right] \\
&= I_1 + I_2,
\end{aligned}$$

with $I_1 = \mathbb{E}\left[\mathrm{e}^{-ua_n(n-1-\frac{\tau_n}{\gamma})} \left(1 - \mathrm{e}^{-u^\alpha a_n^\alpha (\frac{\tau_n}{\gamma} - n)}\right)\right]$ and $I_2 = \mathbb{E}\left[M_{ua_n,\tau_n}\, \mathrm{e}^{u^\alpha + \sum_{i=1}^{\tau_n} R(Y_{i-1}^{(n)}, ua_n)}\right]$.

Using (25) and $\mathbb{E}[M_{ua_n,\tau_n}] = 1$, we get

$$\mathrm{e}^{u^\alpha - C} 25^{(u)n^{-\varepsilon_1}} \leq I_2 \leq \mathrm{e}^{u^\alpha + C} 25^{(u)n^{-\varepsilon_1}} .$$

This implies that $\lim_{n\to\infty} I_2 = \mathrm{e}^{u^\alpha}$.

We now prove that $\lim_{n\to\infty} I_1 = 0$. Recall that $\tau_n \leq n-1$ so that $\tau_n a_n^\alpha \leq 1$ and thanks to (25), we get

$$\mathbb{E}[\mathrm{e}^{-ua_n(n-1-\frac{\tau_n}{\gamma})}] = \mathbb{E}\left[M_{ua_n,\tau_n}\, \mathrm{e}^{\frac{u^\alpha \tau_n a_n^\alpha}{\gamma} + \sum_{i=1}^{\tau_n} R(Y_{i-1}^{(n)}, ua_n)}\right] \leq M(u)\mathbb{E}[M_{ua_n,\tau_n}] = M(u),$$

where $M(u)$ is a constant which does not depend on n. By Cauchy-Schwarz' inequality, we get that

$$I_1 = \mathbb{E}\left[\mathrm{e}^{-ua_n(n-1-\frac{\tau_n}{\gamma})}\left(1-\mathrm{e}^{-u^\alpha a_n^\alpha(\frac{\tau_n}{\gamma}-n)}\right)\right]^2 \leq \mathbb{E}\left[\mathrm{e}^{-2ua_n(n-1-\frac{\tau_n}{\gamma})}\right]\mathbb{E}\left[\left(1-\mathrm{e}^{-u^\alpha a_n^\alpha(\frac{\tau_n}{\gamma}-n)}\right)^2\right]$$
$$\leq M(2u)\mathbb{E}\left[\left(1-\mathrm{e}^{-u^\alpha \frac{1}{n}(\frac{\tau_n}{\gamma}-n)}\right)^2\right].$$

Notice $(\frac{1}{n}(\frac{\tau_n}{\gamma}-n), n\geq 1)$ is bounded from below and above by finite constants, and thanks to Lemma 3.3 it converges to 0 in probability. Hence, we deduce that

$$\lim_{n\to\infty}\mathbb{E}\left[\left(1-\mathrm{e}^{-u^\alpha \frac{1}{n}(\frac{\tau_n}{\gamma}-n)}\right)^2\right]=0.$$

This implies that $\lim_{n\to\infty} I_1 = 0$.

From the convergence of I_1 and I_2, we deduce from (26) that $\lim_{n\to\infty}\mathbb{E}\left[\mathrm{e}^{-ua_n(n-1-\frac{\tau_n}{\gamma})}\right]=\mathrm{e}^{u^\alpha}$. This ends the proof of the Proposition. □

We now give a general result.

Proposition 3.4. *We assume that $\zeta > 1-1/\alpha$. Let $f_n : \mathbb{R}_+ \to \mathbb{R}_+$ be uniformly bounded functions such that*

$$\kappa = \lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{\lfloor n\gamma\rfloor} f_n(k/n)^\alpha$$

exists. Then we have the following convergence in distribution

$$V^{(n)}(f_n) := n^{-\frac{1}{\alpha}}\sum_{k=1}^{\tau_n} f_n(k/n)(X_k^n - \frac{1}{\gamma}) \xrightarrow[n\to\infty]{(d)} \kappa^{1/\alpha}V_1. \tag{27}$$

In particular, if $f : \mathbb{R}_+ \to \mathbb{R}_+$ is a bounded locally Riemann integrable function, then

$$V^{(n)}(f) = n^{-\frac{1}{\alpha}}\sum_{k=1}^{\tau_n} f(k/n)(X_k^n - \frac{1}{\gamma}) \xrightarrow[n\to\infty]{(d)} \int_0^\gamma f(t)dV_t, \tag{28}$$

where the distribution of $\int_0^\gamma f(t)dV_t$ is characterized by its Laplace transform: for $u\geq 0$,

$$\mathbb{E}[\exp(-u\int_0^\gamma f(t)dV_t)] = \exp\left(\frac{u^\alpha}{\gamma}\int_0^\gamma f^\alpha(t)\,dt\right). \tag{29}$$

If we apply this Proposition with step functions, we deduce the following result.

Corollary 3.5. *We assume that $\zeta > 1-1/\alpha$. Let $V_t^{(n)} = V^{(n)}(\mathbf{1}_{[0,t]}) = n^{-1/\alpha}\sum_{k=1}^{\lfloor nt\rfloor\wedge\tau_n}(X_k^{(n)} - \frac{1}{\gamma})$ for $t\in[0,\gamma)$, and $V_\gamma^{(n)} = V^{(n)}(\mathbf{1}) = n^{-1/\alpha}\left(n-1-\frac{\tau_n}{\gamma}\right)$. The finite-dimensional marginals of the process $(V_t^{(n)}, t\in[0,\gamma])$ converges in law to those of the process $(V_t, t\in[0,\gamma])$.*

Proof. Thanks to [18], it is enough to prove that

$$\mathbb{E}[\exp(-uV^{(n)}(f_n))] \xrightarrow[n\to\infty]{} \mathrm{e}^{\kappa u^\alpha/\gamma}.$$

Taking uf_n as f_n, we shall only consider the case $u=1$.

We set $a = \sup_{n\geq 1, x\geq 0} |f_n(x)|$ and for any bounded function g,

$$A_n(g) = \exp \sum_{k=1}^{\tau_n} \left(-n^{-1/\alpha} g(k/n) X_k^{(n)} - \log \phi_{Y_{k-1}^{(n)}} (n^{-\frac{1}{\alpha}} g(k/n)) \right).$$

A martingale argument provides that $\mathbb{E}[A_n(g)] = 1$. Using (19), we get that :

$$\begin{aligned} A_n(g) &= \exp \left(-n^{-1/\alpha} \sum_{k=1}^{\tau_n} g(k/n)(X_k^{(n)} - \frac{1}{\gamma}) - n^{-1} \sum_{k=1}^{\tau_n} \frac{g^\alpha(k/n)}{\gamma} - \sum_{k=1}^{\tau_n} R(Y_{k-1}^{(n)}, n^{-\frac{1}{\alpha}} g(k/n)) \right) \\ &= \exp \left(-V^{(n)}(g) - n^{-1} \sum_{k=1}^{\tau_n} \frac{g^\alpha(k/n)}{\gamma} - \sum_{k=1}^{\tau_n} R(Y_{k-1}^{(n)}, n^{-\frac{1}{\alpha}} g(k/n)) \right). \end{aligned}$$

Let $\Lambda_n = n^{-1} \sum_{k=1}^{\lfloor n\gamma \rfloor} \frac{f_n^\alpha(k/n)}{\gamma} - n^{-1} \sum_{k=1}^{\tau_n} \frac{f_n^\alpha(k/n)}{\gamma}$ and write

$$\mathbb{E}\left[\mathrm{e}^{-V^{(n)}(f_n)}\right] = I_1 + I_2$$

with $I_1 = \mathbb{E}\left[\mathrm{e}^{-V^{(n)}(f_n)} \left(1 - \mathrm{e}^{\Lambda_n}\right)\right]$ and $I_2 = \mathbb{E}\left[\mathrm{e}^{-V^{(n)}(f_n)} \mathrm{e}^{\Lambda_n}\right]$.

First of all, let us prove that I_1 converges to 0 when n tends to ∞. Recall that the functions f_n are uniformly bounded by a. Thanks to (25), we have

$$\mathbb{E}[\mathrm{e}^{-2V^{(n)}(f_n)}] = \mathbb{E}[\mathrm{e}^{-V^{(n)}(2f_n)}] = \mathbb{E}\left[A_n(2f_n)\, \mathrm{e}^{n^{-1} \sum_{k=1}^{\tau_n} \frac{2^\alpha f_n^\alpha(k/n)}{\gamma} + \sum_{k=1}^{\tau_n} R(Y_{k-1}^{(n)}, n^{-\frac{1}{\alpha}} 2 f_n(k))} \right] \leq M,$$

where M is a finite constant which does not depend on n. By Cauchy-Schwarz' inequality, we get that

$$(I_1)^2 \leq \left(\mathbb{E}\left[\mathrm{e}^{-V^{(n)}(f_n)} \left|1 - \mathrm{e}^{\Lambda_n}\right| \right] \right)^2 \leq \mathbb{E}\left[\mathrm{e}^{-V^{(n)}(2f_n)}\right] \mathbb{E}\left[\left(1 - \mathrm{e}^{\Lambda_n}\right)^2 \right] \leq M \mathbb{E}\left[\left(1 - \mathrm{e}^{\Lambda_n}\right)^2 \right].$$

Moreover as $|1 - \mathrm{e}^x| \leq \mathrm{e}^{|x|} - 1$ and $\Lambda_n \leq \frac{a^\alpha}{n\gamma} |\lfloor n\gamma \rfloor - \tau_n|$, we get

$$\mathbb{E}\left[\left(1 - \mathrm{e}^{\Lambda_n}\right)^2 \right] \leq \mathbb{E}\left[\left(1 - \mathrm{e}^{\frac{|\lfloor n\gamma \rfloor - \tau_n| a^\alpha}{n\gamma}} \right)^2 \right]. \tag{30}$$

The quantity $\frac{|\lfloor n\gamma \rfloor - \tau_n| a^\alpha}{n\gamma}$ is bounded and goes to 0 in probability when n goes to infinity. Therefore, the right-hand side of (30) converges to 0. This implies that $\lim_{n\to\infty} I_1 = 0$.

Let us now consider the convergence of I_2. Remark that

$$I_2 = \mathbb{E}\left[A_n(f_n)\, \mathrm{e}^{n^{-1} \sum_{k=1}^{\lfloor n\gamma \rfloor} \frac{f_n^\alpha(k/n)}{\gamma} + \sum_{k=1}^{\tau_n} R(Y_{k-1}^{(n)}, n^{-\frac{1}{\alpha}} f_n(k))} \right].$$

Recall that f_n is bounded by a and that $\mathbb{E}[A_n(f_n)] = 1$. Using Lemma 3.2, we get for some $\varepsilon > 0$

$$\begin{aligned} &\mathrm{e}^{-C_{25}(a) n^{-\varepsilon_1} - n^{-1} \sum_{k=1}^{\lfloor n\gamma \rfloor} \frac{f_n^\alpha(k/n)}{\gamma}} \\ &\qquad \leq \mathbb{E}\left[A_n(f_n)\, \mathrm{e}^{n^{-1} \sum_{k=1}^{\lfloor n\gamma \rfloor} \frac{f_n^\alpha(k/n)}{\gamma} + \sum_{k=1}^{\tau_n} R(Y_{k-1}^{(n)}, n^{-\frac{1}{\alpha}} f_n(k))} \right] \\ &\qquad\qquad \leq \mathrm{e}^{C_{25}(a) n^{-\varepsilon_1} + n^{-1} \sum_{k=1}^{\lfloor n\gamma \rfloor} \frac{f_n^\alpha(k/n)}{\gamma}}. \end{aligned} \tag{31}$$

As $\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{\lfloor n\gamma \rfloor} f_n^\alpha(k/n) = \kappa$, we get that $\lim_{n\to\infty} I_2 = \mathrm{e}^{\kappa/\gamma}$, which achieves the proof of (27).

To get (28), notice that $\kappa = \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{\lfloor n\gamma \rfloor} f(k/n)^\alpha = \int_0^\gamma f(t)^\alpha \, dt.$ □

4. First approximation of the length of the coalescent tree

Let $\alpha \in (1,2)$. We assume that $\rho(t) = C_0 t^{-\alpha} + O(t^{-\alpha+\zeta})$ for some $C_0 > 0$ and $\zeta > 1 - 1/\alpha$.

Recall that the length of the coalescent tree up to the $\lfloor nt \rfloor$-th coalescence is, for $t \geq 0$, given by (3). The next Lemma gives an upper bound on the error when one replaces the exponential random variables by their mean.

Lemma 4.1. *For $t \geq 0$, let*

$$\tilde{L}_t^{(n)} = \sum_{k=0}^{\lfloor nt \rfloor \wedge (\tau_n - 1)} \frac{Y_k^{(n)}}{g_{Y_k^{(n)}}}.$$

There exists a finite constant C_{32} such that for all $t \geq 0$, we have

$$\mathbb{E}\left[(L_t^{(n)} - \tilde{L}_t^{(n)})^2 \right] \leq C_{32} \begin{cases} n^{3-2\alpha} & \text{if } \alpha < 3/2, \\ \log(n) & \text{if } \alpha = 3/2, \\ 1 & \text{if } \alpha > 3/2. \end{cases} \tag{32}$$

Proof. Conditionally on $\mathcal{Y}$, the random variables $\frac{Y_k^{(n)}}{g_{Y_k^{(n)}}}(E_k - 1)$ are independent with zero mean. We deduce that

$$\begin{aligned} \mathbb{E}\left[(L_t^{(n)} - \tilde{L}_t^{(n)})^2 | \mathcal{Y} \right] &= \mathbb{E}\left[\left(\sum_{k=0}^{\lfloor nt \rfloor \wedge (\tau_n - 1)} \frac{Y_k^{(n)}}{g_{Y_k^{(n)}}} (E_k - 1) \right)^2 | \mathcal{Y} \right] \\ &= \sum_{k=0}^{\lfloor nt \rfloor \wedge (\tau_n - 1)} \left(\frac{Y_k^{(n)}}{g_{Y_k^{(n)}}} \right)^2 \\ &\leq \sum_{\ell=1}^{n} \left(\frac{\ell}{g_\ell} \right)^2 . \end{aligned}$$

Thanks to (13), we get

$$\mathbb{E}\left[(L_t^n - \tilde{L}_t^{(n)})^2 | \mathcal{Y} \right] \leq c \sum_{\ell=1}^{n} \ell^{2-2\alpha} \leq c \begin{cases} n^{3-2\alpha} & \text{if } \alpha < 3/2, \\ \log(n) & \text{if } \alpha = 3/2, \\ 1 & \text{if } \alpha > 3/2, \end{cases}$$

where c is non random. This implies the result. □

Lemma 4.2. *For $t \geq 0$, let*

$$\hat{L}_t^{(n)} = \sum_{k=0}^{\lfloor nt \rfloor \wedge (\tau_n - 1)} \left(Y_k^{(n)} \right)^{-\gamma} .$$

There exists a finite constant C_{33} such that for all $t \geq 0$, we have

$$|\tilde{L}_t^{(n)} - \frac{\hat{L}_t^{(n)}}{C_0\Gamma(2-\alpha)}| \leq C_{33} \begin{cases} n^{2-\alpha-\zeta} & \text{if } \zeta < 2-\alpha, \\ \log(n) & \text{if } \zeta = 2-\alpha, \\ 1 & \text{if } \zeta > 2-\alpha. \end{cases} \tag{33}$$

Proof. Use (13) to get that

$$\tilde{L}_t^{(n)} - \frac{\hat{L}_t^{(n)}}{C_0\Gamma(2-\alpha)} = \sum_{k=0}^{\lfloor nt \rfloor \wedge (\tau_n - 1)} \left(Y_k^{(n)}\right)^{-\gamma} O\left(\left(Y_k^{(n)}\right)^{-\min(\zeta,1)}\right).$$

We deduce that

$$|\tilde{L}_t^{(n)} - \frac{\hat{L}_t^{(n)}}{C_0\Gamma(2-\alpha)}| \leq c \sum_{\ell=1}^{n} \ell^{-\alpha+1-\min(\zeta,1)} \leq c \begin{cases} n^{2-\alpha-\zeta} & \text{if } \zeta < 2-\alpha, \\ \log(n) & \text{if } \zeta = 2-\alpha, \\ 1 & \text{if } \zeta > 2-\alpha. \end{cases}$$

□

5. Limit distribution of $\hat{L}_t^{(n)}$

Let $\alpha \in (1,2)$ and $\gamma = \alpha - 1$. For $t \in [0,\gamma]$, we set

$$v(t) = \int_0^t \left(1 - \frac{r}{\gamma}\right)^{-\gamma} dr.$$

Theorem 5.1. *We assume that $\rho(t) = C_0 t^{-\alpha} + O(t^{-\alpha+\zeta})$ for some $C_0 > 0$ and $\zeta > 1 - 1/\alpha$. Then for all $t \in (0,\gamma)$, we have that*

(1) *The following convergence in probability holds:*

$$n^{-2+\alpha} \hat{L}_t^{(n)} \xrightarrow[n\to\infty]{\mathbb{P}} v(t). \tag{34}$$

(2) *The following convergence in distribution holds:*

$$n^{-1+\alpha-1/\alpha}(\hat{L}_t^{(n)} - n^{2-\alpha} v(t)) \xrightarrow[n\to\infty]{(d)} (\alpha-1)\int_0^t dr\, (1-\frac{r}{\gamma})^{-\alpha} V_r. \tag{35}$$

Proof of Theorem 5.1. Let $\varepsilon_2 \in (0,\gamma)$ and $t \in (0, \gamma - \varepsilon_2)$. We use a Taylor expansion to get

$$\begin{aligned} \hat{L}_t^{(n)} &= \sum_{k=0}^{\lfloor nt \rfloor \wedge (\tau_n - 1)} \left(n - \sum_{i=1}^{k} X_i^{(n)}\right)^{-\gamma} \\ &= \sum_{k=0}^{\lfloor nt \rfloor \wedge (\tau_n - 1)} \left(n - \frac{k}{\gamma} - \sum_{i=1}^{k} (X_i^{(n)} - \frac{1}{\gamma})\right)^{-\gamma} \\ &= \sum_{k=0}^{\lfloor nt \rfloor \wedge (\tau_n - 1)} \left(n - \frac{k}{\gamma}\right)^{-\gamma} (1 - \Delta_{n,k})^{-\gamma} \\ &= I_n + \gamma J_n + \gamma(\gamma+1) R_n \end{aligned} \tag{36}$$

with $\Delta_{n,k} = \dfrac{\sum_{i=1}^{k}(X_i^{(n)} - \frac{1}{\gamma})}{n - k/\gamma}$ and

$$\begin{aligned} I_n &= \sum_{k=0}^{\lfloor nt\rfloor\wedge(\tau_n-1)} \left(n - \frac{k}{\gamma}\right)^{-\gamma}, \\ J_n &= \sum_{k=1}^{\lfloor nt\rfloor\wedge(\tau_n-1)} \left(n - \frac{k}{\gamma}\right)^{-\gamma-1} \sum_{i=1}^{k}(X_i^{(n)} - \frac{1}{\gamma}), \\ R_n &= \sum_{k=1}^{\lfloor nt\rfloor\wedge(\tau_n-1)} \left(n - \frac{k}{\gamma}\right)^{-\gamma} \int_0^{\Delta_{n,k}} (\Delta_{n,k} - t)\,(1-t)^{-\gamma-2}\,dt. \end{aligned}$$

Notice that a.s. $\Delta_{n,k} < 1$, so that R_n is well defined.

Convergence of I_n. We first give an expansion of I_n by considering $I_n = n^{2-\alpha} I_{n,1}\mathbf{1}_{\{nt<\tau_n\}} + I_n\mathbf{1}_{\{nt\geq\tau_n\}}$ with $I_{n,1} = \dfrac{1}{n}\displaystyle\sum_{k=0}^{\lfloor nt\rfloor}\left(1 - \frac{k}{n\gamma}\right)^{-\gamma}$. Standard computation yields

$$I_{n,1} = v(t) + \frac{1}{n}\,h_3(n,t),$$

where $\sup_{t\in(0,\gamma-\varepsilon),n\geq1} |h_3(n,t)| < \infty$. By decomposing according to $\{nt < \tau_n\}$ and $\{nt \geq \tau_n\}$, we deduce that,

$$\mathbb{P}\left(n^{-1+\alpha-1/\alpha}\left|I_n - n^{2-\alpha}v(t)\right| \geq \varepsilon\right) \leq \mathbb{P}(n^{-1/\alpha}|h_3(n,t)| \geq \varepsilon/2) + \mathbb{P}(nt \geq \tau_n).$$

According to Lemma 3.3, τ_n/n converges in probability to $\gamma > t$. This implies that

$$\lim_{n\to\infty} \mathbb{P}(nt \geq \tau_n) = 0. \tag{37}$$

As $n^{-1/\alpha}|h_3(n,t)| \leq \varepsilon$ for n large enough, we deduce the following convergence in probability:

$$n^{-1+\alpha-1/\alpha}\left(I_n - n^{2-\alpha}v(t)\right) \xrightarrow[n\to\infty]{\mathbb{P}} 0. \tag{38}$$

Convergence of J_n. To get the convergence of J_n, notice that

$$J_n = \sum_{i=1}^{\lfloor nt\rfloor\wedge(\tau_n-1)} (X_i^{(n)} - \frac{1}{\gamma}) \sum_{k=i}^{\lfloor nt\rfloor\wedge(\tau_n-1)} \left(n - \frac{k}{\gamma}\right)^{-\alpha} = n^{1-\alpha} J_{n,1}\mathbf{1}_{\{nt<\tau_n\}} + J_n\mathbf{1}_{\{nt\geq\tau_n\}}, \tag{39}$$

with $J_{n,1} = \displaystyle\sum_{i=1}^{\lfloor nt\rfloor\wedge(\tau_n-1)} f_n(i)(X_i^{(n)} - \frac{1}{\gamma})$ and $f_n(r) = \dfrac{1}{n}\displaystyle\sum_{j=\lfloor nr\rfloor}^{\lfloor nt\rfloor}\left(1 - \frac{j}{n\gamma}\right)^{-\alpha}$. The functions f_n are finite and uniformly bounded as for $n \geq 2/\varepsilon_2$,

$$0 \leq f_n(r) \leq f_n(0) = \frac{1}{n}\sum_{k=0}^{\lfloor nt\rfloor}\left(1 - \frac{k}{n\gamma}\right)^{-\alpha} \leq \int_0^{\gamma-\varepsilon_2/2}\left(1 - \frac{s}{\gamma}\right)^{-\alpha} ds < \infty.$$

Notice that

$$\kappa = \lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{\lfloor n\gamma\rfloor} f_n(k)^\alpha = \int_0^t dr\left(\int_r^t (1 - \frac{s}{\gamma})^{-\alpha}\,ds\right)^\alpha.$$

We deduce from Proposition 3.4 that $(n^{-\frac{1}{\alpha}} J_{n,1}, n \geq 2)$ converges in distribution to $\kappa^{1/\alpha} V_1$. For $\varepsilon' > 0$, we have $\mathbb{P}(\mathbf{1}_{\{nt \geq \tau_n\}} |J_n| \geq \varepsilon') \leq \mathbb{P}(nt \geq \tau_n)$. Then we use (39) and (37) to conclude that the following convergence in distribution holds:

$$n^{-1+\alpha-1/\alpha} J_n \xrightarrow[n\to\infty]{\text{(d)}} \kappa^{1/\alpha} V_1. \tag{40}$$

Convergence of R_n. We shall now prove that $n^{-1+\alpha-1/\alpha} R_n$ converges to 0 in probability. Let $\varepsilon \in (0, \gamma)$. We have $R_n = R_{n,1} + R_{n,2}$, with

$$\begin{aligned}
R_{n,1} &= \sum_{k=1}^{\lfloor nt \rfloor} \left(n - \frac{k}{\gamma}\right)^{-\gamma} \mathbf{1}_{\{k<\tau_n\}} R_{n,1,k}, \\
R_{n,1,k} &= \mathbf{1}_{\{\Delta_{n,k}<1-\varepsilon\}} \int_0^{\Delta_{n,k}} (\Delta_{n,k} - t)\,(1-t)^{-\gamma-2}\, dt, \\
R_{n,2} &= \sum_{k=1}^{\lfloor nt \rfloor} \left(n - \frac{k}{\gamma}\right)^{-\gamma} \mathbf{1}_{\{k<\tau_n\}} \mathbf{1}_{\{\Delta_{n,k}\geq 1-\varepsilon\}} \int_0^{\Delta_{n,k}} (\Delta_{n,k} - t)\,(1-t)^{-\gamma-2}\, dt.
\end{aligned}$$

We have for $k \leq n(\gamma - \varepsilon_2)$,

$$\mathbb{E}[|R_{n,1,k}|] \leq c\, \mathbb{E}[(\Delta_{n,k})^2] \leq \frac{c}{n^2}\, \mathbb{E}\left[\left(\sum_{i=1}^{k} (X_i^{(n)} - \frac{1}{\gamma})\right)^2\right].$$

Recall $\mathcal{Y} = (\mathcal{Y}_k, k \geq 0)$ is the filtration generated by Y. We consider the $\mathcal{Y}$-martingale $N_r = \sum_{j=1}^{r} \Delta N_r$, with $\Delta N_r = X_r^{(n)} - \mathbb{E}[X_r^{(n)}|\mathcal{Y}_{r-1}]$. We have

$$\mathbb{E}\left[\left(\sum_{i=1}^{k} (X_i^{(n)} - \frac{1}{\gamma})\right)^2\right] \leq 2\mathbb{E}\left[N_k^2\right] + 2\mathbb{E}\left[\left(\sum_{i=1}^{k} (\mathbb{E}[X_i^{(n)}|\mathcal{Y}_{i-1}] - \frac{1}{\gamma})\right)^2\right].$$

Notice that

$$\mathbb{E}\left[N_k^2\right] = \mathbb{E}\left[\sum_{i=1}^{k} (\Delta N_i)^2\right] \leq \mathbb{E}\left[\sum_{i=1}^{k} \mathbb{E}[(X_i^{(n)})^2|\mathcal{Y}_{i-1}]\right] \leq \mathbb{E}\left[\sum_{i=1}^{k} (X_i^{(n)})^2\right].$$

Using that, conditionally on $\mathcal{Y}_{i-1}$, $X_i^{(n)}$ and $X_1^{(Y_{i-1})}$ have the same distribution, we get that

$$\mathbb{E}\left[N_k^2\right] \leq \sum_{j=1}^{n} \mathbb{E}[(X_1^{(j)})^2].$$

Thanks to (18) and (13), we deduce that

$$\mathbb{E}\left[N_k^2\right] \leq C_{18} \sum_{j=1}^{n} \frac{j^2}{g_j} \leq c \sum_{j=1}^{n} j^{2-\alpha} \leq c\, n^{3-\alpha}.$$

Using (15) and (13), we get

$$\begin{aligned}\mathbb{E}\left[\left(\sum_{i=1}^{k}(\mathbb{E}[X_i^{(n)}|\mathcal{Y}_{i-1}]-\frac{1}{\gamma})\right)^2\right] &\leq \mathbb{E}\left[\left(\sum_{i=1}^{k}|\mathbb{E}[X_i^{(n)}|\mathcal{Y}_{i-1}]-\frac{1}{\gamma}|\right)^2\right]\\ &\leq \mathbb{E}\left[\left(\sum_{i=1}^{k}C_{15}\varphi_{Y_{i-1}}\right)^2\right]\\ &\leq c\left(\sum_{j=1}^{n}\varphi_j\right)^2 \leq c\, n^{3-\alpha},\end{aligned}$$

where for the last inequality we used (14) with $\varepsilon_0>0$ small enough (such that $1+2\varepsilon_0<\alpha$) and the fact that $\zeta>1-1/\alpha$ implies $2-2\zeta\leq 3-\alpha$ as $\alpha\in(1,2)$. This implies that $\mathbb{E}[|R_{n,1,k}|]\leq c\, n^{1-\alpha}$ and therefore $\mathbb{E}[|R_{n,1}|]\leq c\, n^{3-2\alpha}$. In particular, we get that $(n^{-1+\alpha-1/\alpha}R_{n,1}, n\geq 1)$ converges in probability to 0 since $-1+\alpha-1/\alpha+3-2\alpha=-(\alpha-1)^2/\alpha<0$ for $\alpha>1$.

We now consider $R_{n,2}$. Suppose that $k\leq\lfloor nt\rfloor-1$ satisfies $\Delta_{n,k}\geq 1-\varepsilon$ on $\{nt<\tau_n\}$. Then on $\{nt<\tau_n\}$, we have

$$\Delta_{n,k+1}=\Delta_{n,k}+\frac{X_{k+1}^{(n)}-\frac{1}{\gamma}+\frac{\Delta_{n,k}}{\gamma}}{n-(k+1)/\gamma}\geq\Delta_{n,k}+\frac{X_{k+1}^{(n)}-\frac{\varepsilon}{\gamma}}{n-(k+1)/\gamma}\geq\Delta_{n,k},$$

where we used that $\gamma>\varepsilon$ for the first inequality and $X_{k+1}^{(n)}\geq 1$ for the last. In particular, on $\{nt<\tau_n\}$, if $\Delta_{n,k}\geq 1-\varepsilon$ for some $k\leq\lfloor nt\rfloor$, then we have $\Delta_{n,\lfloor nt\rfloor}\geq 1-\varepsilon$. This implies that $\mathbf{1}_{\{nt<\tau_n\}}R_{n,2}=\mathbf{1}_{\{\Delta_{n,\lfloor nt\rfloor}\geq 1-\varepsilon\}}\mathbf{1}_{\{nt<\tau_n\}}R_{n,2}$. With the notations of Corollary 3.5, we have

$$\{nt<\tau_n\}\cap\{\Delta_{n,\lfloor nt\rfloor}\geq 1-\varepsilon\}\subset\{V_t^{(n)}\geq(1-\varepsilon)(n-\frac{\lfloor nt\rfloor}{\gamma})n^{-1/\alpha}\}\subset\{n^{-1+1/\alpha}V_t^{(n)}\geq c\},$$

and then for any $\varepsilon'>0$

$$\begin{aligned}\mathbb{P}(n^{-1+\alpha-1/\alpha}|R_{n,2}|\geq\varepsilon',\ nt<\tau_n)&=\mathbb{P}(\mathbf{1}_{\{\Delta_{n,\lfloor nt\rfloor}\geq 1-\varepsilon\}}n^{-1+\alpha-1/\alpha}|R_{n,2}|\geq\varepsilon',\ nt<\tau_n)\\ &\leq\mathbb{P}(\Delta_{n,\lfloor nt\rfloor}\geq 1-\varepsilon,,\ nt<\tau_n)\\ &\leq\mathbb{P}(n^{-1+1/\alpha}V_t^{(n)}\geq c).\end{aligned}$$

Use the convergence of $V_t^{(n)}$, see Corollary 3.5, to get that the right-hand side of the last inequality converges to 0 as n goes to infinity. Then notice that $\mathbb{P}(n^{-1+\alpha-1/\alpha}|R_{n,2}|\geq\varepsilon',\ nt\geq\tau_n)\leq\mathbb{P}(nt\geq\tau_n)$ which converges to 0 thanks to (37).

Thus the following convergence in probability holds:

$$n^{-1+\alpha-1/\alpha}R_n \xrightarrow[n\to\infty]{\mathbb{P}} 0. \tag{41}$$

We deduce from (36), (38), (40) and (41) that

$$n^{-1+\alpha-1/\alpha}\left(\hat{L}_t^{(n)}-n^{2-\alpha}v(t)\right)\xrightarrow[n\to\infty]{\text{(d)}}\gamma\left[\int_0^t dr\,\left(\int_r^t(1-\frac{s}{\gamma})^{-\alpha}ds\right)^\alpha\right]^{1/\alpha}V_1. \tag{42}$$

To conclude, use (29) to get that $\gamma\left[\int_0^t dr\left(\int_r^t(1-\frac{s}{\gamma})^{-\alpha}ds\right)^\alpha\right]^{1/\alpha}V_1$ is distributed as $\gamma\int_0^t dV_r\int_r^t(1-\frac{s}{\gamma})^{-\alpha}ds$ which in turn is equal to $\int_0^t dr\,(1-\frac{r}{\gamma})^{-\alpha}V_r$.

□

6. Proof of the main result

Let $\alpha_0 = \dfrac{1+\sqrt{5}}{2}$. Notice that for $\alpha \in (1, \alpha_0)$, we have $-1+\alpha-1/\alpha < 0$, whereas for $\alpha \geq \alpha_0$, $-1+\alpha-1/\alpha \geq 0$. Recall $\gamma = \alpha - 1$. We define $a(t)$ for $t \in [0, \gamma]$ by

$$a(t) = \frac{v(t)}{C_0\Gamma(2-\alpha)}, \quad \text{where} \quad v(t) = \int_0^t \left(1 - \frac{r}{\gamma}\right)^{-\gamma} dr.$$

We also set $V_t^* = \dfrac{\alpha-1}{C_0\Gamma(2-\alpha)} \displaystyle\int_0^t (1-\frac{r}{\gamma})^{-\alpha} V_r \, dr$ for $t \in (0, \gamma)$.

Theorem 6.1. *We assume that $\rho(t) = C_0 t^{-\alpha} + O(t^{-\alpha+\zeta})$ for some $C_0 > 0$ and $\zeta > 1 - 1/\alpha$. Then for all $t \in (0, \gamma)$, we have that*

(1) *The following convergence in probability holds:*

$$n^{-2+\alpha} L_t^{(n)} \xrightarrow[n\to\infty]{\mathbb{P}} a(t). \tag{43}$$

(2) *If $\alpha \in (1, \alpha_0)$, the following convergence in distribution holds:*

$$n^{-1+\alpha-1/\alpha}\left(L_t^{(n)} - a(t)n^{2-\alpha}\right) \xrightarrow[n\to\infty]{(d)} V_t^*. \tag{44}$$

(3) *If $\alpha \in [\alpha_0, 2)$, the following convergence in probability holds: If $\varepsilon > 0$,*

$$n^{-\varepsilon}\left(L_t^{(n)} - a(t)n^{2-\alpha}\right) \xrightarrow[n\to\infty]{\mathbb{P}} 0. \tag{45}$$

Proof. First of all, let us consider the case $\alpha \in (1, \alpha_0)$. Lemma 4.1 and Tchebychev inequality imply that for $\alpha \in (1, \alpha_0)$, we have the following convergence in probability

$$\lim_{n\to\infty} n^{-1+\alpha-1/\alpha} |L_t^{(n)} - \tilde{L}_t^{(n)}| = 0.$$

This and Lemma 4.2 imply that for $\alpha \in (1, \alpha_0)$, we have the following convergence in probability

$$\lim_{n\to\infty} n^{-1+\alpha-1/\alpha} |L_t^{(n)} - \frac{\hat{L}_t^{(n)}}{C_0\Gamma(2-\alpha)}| = 0.$$

The result is then a direct consequence of Theorem 5.1.

For $\alpha \in [\alpha_0, 2)$, note that $\alpha > 3/2$ and $-1+\alpha-1/\alpha \geq 0$. As $\zeta > 1-1/\alpha$ and $\alpha > \alpha_0$ i.e. $1-1/\alpha > 2-\alpha$, we get $\zeta > 2-\alpha$. We then use Lemma 4.1, Lemma 4.2 (only with $\zeta > 2-\alpha$) and Theorem 5.1 to get (45), and then (43). □

Let $K_t^{(n)}$ be the number of mutations up to the $\lfloor nt \rfloor$-th coalescence, for $t \in (0, \gamma)$. conditionally on $L_t^{(n)}$, $K_t^{(n)}$ is a Poisson r.v. with parameter $\theta L_t^{(n)}$. The next Corollary is a consequence of Theorem 6.1.

Corollary 6.2. *We assume that $\rho(t) = C_0 t^{-\alpha} + O(t^{-\alpha+\zeta})$ for some $C_0 > 0$ and $\zeta > 1 - 1/\alpha$. Let $t \in (0, \gamma)$ and G be a standard Gaussian r.v., independent of V.*

(1) *For $\alpha \in (1, \sqrt{2})$, we have*

$$n^{-1+\alpha-1/\alpha}(K_t^{(n)} - \theta a(t) n^{2-\alpha}) \xrightarrow[n\to\infty]{(d)} \theta V_t^*.$$

(2) *For* $\alpha \in (\sqrt{2}, 2)$, *we have*

$$n^{-1+\alpha/2}(K_t^{(n)} - \theta a(t) n^{2-\alpha}) \xrightarrow[n\to\infty]{(d)} \sqrt{\theta a(t)} G.$$

(3) *For* $\alpha = \sqrt{2}$, *we have* $-1 + \alpha - \frac{1}{\alpha} = 1 - \frac{\alpha}{2}$ *and*

$$n^{-1+\alpha-1/\alpha}(K_t^{(n)} - \theta a(t) n^{2-\alpha}) \xrightarrow[n\to\infty]{(d)} \theta V_t^* + \sqrt{\theta a(t)} G.$$

Proof. Let us compute the characteristic function $\psi_n(u,v)$ of the 2-dimensional r.v. (G_n, H_n) with

$$G_n = \frac{K_t^{(n)} - \theta L_t^{(n)}}{\sqrt{\theta a(t) n^{2-\alpha}}} \quad \text{and} \quad H_n = n^{-1+\alpha-1/\alpha}\left(L_t^{(n)} - a(t)n^{2-\alpha}\right).$$

Using that, conditionally on $L_t^{(n)}$, the law of $K_t^{(n)}$ is a Poisson distribution with parameter $\theta L_t^{(n)}$, we have

$$\psi_n(u,v) = \mathbb{E}\left[\mathrm{e}^{iuG_n}\,\mathrm{e}^{ivH_n}\right] = \mathbb{E}\left[\mathrm{e}^{-\theta L_t^{(n)}\left(1-\mathrm{e}^{iu/\sqrt{\theta a(t)n^{2-\alpha}}} + iu/\sqrt{\theta a(t) n^{2-\alpha}}\right)}\,\mathrm{e}^{ivH_n}\right].$$

We first consider the case $\alpha \in (1, \alpha_0)$. Using Theorem 6.1, we get that

$$-\theta L_t^{(n)}\left(1 - \mathrm{e}^{iu/\sqrt{\theta a(t) n^{2-\alpha}}} + iu/\sqrt{\theta a(t) n^{2-\alpha}}\right)$$

tends to $-u^2/2$ in probability and has a non-negative real part. Hence, applying Theorem 6.1 again, we get that (G_n, H_n) converges in distribution to (G, V_t^*), where G is a standard Gaussian r.v. independent of V. Notice that

$$K_t^{(n)} = \theta a(t) n^{2-\alpha} + \theta n^{1-\alpha+1/\alpha} H_n + \sqrt{\theta a(t)}\, n^{1-\alpha/2} G_n.$$

We have $\sqrt{2} < \alpha_0$. To conclude when $\alpha < \alpha_0$, use that $1 - \alpha + 1/\alpha$ is smaller (resp. equal to) $1 - \alpha/2$ if and only if $\alpha > \sqrt{2}$ (resp. $\alpha = \sqrt{2}$).

Now we consider $\alpha \in [\alpha_0, 2)$. We write

$$n^{-1+\alpha/2}(K_t^{(n)} - \theta a(t) n^{2-\alpha}) = \sqrt{\theta a(t)} G_n + n^{-1+\alpha/2}(L_t^{(n)} - a(t) n^{2-\alpha}).$$

Using Theorem 6.1, we still get that G_n converges in law to G. Moreover, (45) implies that $n^{-1+\alpha/2}(L_t^{(n)} - a(t)n^{2-\alpha})$ converges to 0 in probability. This gives the result. □

References

[1] J. Berestycki, N. Berestycki, and J. Schweinsberg. Beta-coalescents and continuous stable random trees. *Ann. Probab.*, To appear.

[2] J. Berestycki, N. Berestycki, and J. Schweinsberg. Small time properties of beta-coalescents. *Ann. Inst. H. Poincaré Probab. Statist.*, To appear.

[3] J. Bertoin. *Lévy processes*. Cambridge University Press, Cambridge, 1996.

[4] J. Bertoin and J.-F. Le Gall. Stochastic flows associated to coalescent processes. III. Limit theorems. *Illinois J. Math.*, 50(1-4):147–181 (electronic), 2006.

[5] M. Birkner, J. Blath, M. Capaldo, A. Etheridge, M. Möhle, J. Schweinsberg, and A. Wakolbinger. Alpha-stable branching and beta-coalescents. *Electron. J. Probab.*, 10:no. 9, 303–325 (electronic), 2005.

[6] E. Bolthausen and A.-S. Sznitman. On Ruelle's probability cascades and an abstract cavity method. *Comm. Math. Phys.*, 197(2):247–276, 1998.

[7] E. G. Boom, J. D. G.and Boulding and A. T. Beckenbach. Mitochondrial dna variation in introduced populations of pacific oyster, crassostrea gigas, in british columbia. *Can. J. Fish. Aquat. Sci.*, 51:1608–1614, 1994.

[8] L. Breiman. *Probability*, volume 7 of *Classics in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1992. Corrected reprint of the 1968 original.

[9] M. Drmota, A. Iksanov, M. Möhle, and U. Rösler. Asymptotic results concerning the total branch length of the Bolthausen-Sznitman coalescent. *Stoch. Process. Appl.*, 117, To appear.
[10] B. Eldon and J. Wakeley. Coalescent processes when the distribution of offspring number among individuals is highly skewed. *Genetics*, 172:2621–2633, 2006.
[11] W. Feller. *An introduction to probability theory and its applications*, volume II. John WILEY & Sons, 1971.
[12] A. Gnedin and Y. Yakubovich. On the number of collisions in Λ-coalescents, 2007.
[13] A. Iksanov and M. Möhle. On a random recursion related to absorption times of death Markov chains, 2007.
[14] O. Kallenberg. *Foundations of modern probability*. Probability and its Applications (New York). Springer-Verlag, New York, second edition, 2002.
[15] J. F. C. Kingman. The coalescent. *Stochastic Process. Appl.*, 13(3):235–248, 1982.
[16] J. F. C. Kingman. Origins of the coalescent 1974–1982. *Genetics*, 156:1461–1463, 2000.
[17] M. Möhle. On the number of segregating sites for populations with large family sizes. *Adv. in Appl. Probab.*, 38(3):750–767, 2006.
[18] A. Mukherjea, M. Rao, and S. Suen. A note on moment generating functions. *Statist. Probab. Lett.*, 76(11):1185–1189, 2006.
[19] J. Pitman. Coalescents with multiple collisions. *Ann. Probab.*, 27(4):1870–1902, 1999.
[20] S. Sagitov. The general coalescent with asynchronous mergers of ancestral lines. *J. Appl. Probab.*, 36(4):1116–1125, 1999.
[21] J. Schweinsberg. Coalescent processes obtained from supercritical Galton-Watson processes. *Stochastic Process. Appl.*, 106(1):107–139, 2003.
[22] G. A. Watterson. On the number of segregating sites in genetical models without recombination. *Theoret. Population Biology*, 7:256–276, 1975.

CERMICS, École des Ponts, ParisTech, 6-8 av. Blaise Pascal, Champs-sur-Marne, 77455 Marne La Vallée, France.
E-mail address: delmas@cermics.enpc.fr

UFR de Mathématiques et d'Informatique, Université René Descartes, 45 rue des Saints Pères, 75270 Paris Cedex 06, France
E-mail address: dhersin@math-info.univ-paris5.fr

UFR de Mathématiques et d'Informatique, Université René Descartes, 45 rue des Saints Pères, 75270 Paris Cedex 06, France
E-mail address: Arno.Jegousse@math-info.univ-paris5.fr